Connectionist Approaches to Understanding Aphasic Perseveration

Stephen J. Gotts, Ph.D.,[1,3,4] **and David C. Plaut, Ph.D.**[1–3]

ABSTRACT

Aphasic patients make a variety of speech errors, including perseverations, in tasks that involve a linguistic component. What do perseverative and other errors imply about the nature of the neurologically damaged and intact language systems? Here we discuss the insights into the mechanisms of aphasic perseveration afforded by connectionist models. As a base for discussion, we review the Plaut and Shallice model of optic aphasic errors in object naming, which relies primarily on short-term learning mechanisms to produce perseverations. We then point out limitations of the model in addressing more recent data collected on aphasic perseveration and explain how incorporating information about the interaction of neuromodulatory systems and learning in the brain may help to overcome these limitations.

KEYWORDS: Aphasia, connectionist, neuromodulation, perseveration, priming

Learning Outcomes: As a result of this activity, the participant will be able to (1) identify the mechanistic principles of connectionist models that lead to recurrent perseverations; (2) characterize how these principles differ from those that produce other types of errors, such as ''visual'' and ''semantic''; and (3) describe limitations of the current principles and how they might be modified to incorporate neuroscientific findings on neuromodulation and learning.

After stroke or brain injury, aphasic patients commonly exhibit a range of errors in spontaneous speech and in tasks requiring a verbal response. One of the most intriguing error types for language researchers is *perseveration*—the inappropriate repetition or continuation of a previous utterance or response when a different response is expected.[1] Aphasic perseverations often differ in character from those elicited by patients with other types of deficits,

such as frontal-lobe executive dysfunction, in that they can occur after several correct intervening utterances or responses, leading them to be labeled *recurrent* as opposed to stuck-in-set or continuous.[2,3] Although recurrent perseverations in aphasia may occur after intervening responses, empirical studies have shown that they are most common after little or no delay and attenuate gradually in likelihood over subsequent trials.[4,5] Recurrent perseverations can be on whole words, partial words, or even on parts of drawings,[3,6] and they can be influenced by several stimulus factors, including word length,[7] lexical frequency,[5,8] relationship to the target stimulus,[3,9,10] stimulus repetition,[5] and presentation rate.[8] However, not all of these factors necessarily affect all patients in all behavioral circumstances, and which factors influence performance for any given patient may depend on the particular locus of impairment in the cognitive system, as well as on the particular tasks employed.[4,5,10–12]

These observations raise a couple of fundamental questions: What does the occurrence of perseverations in aphasia tell us about the nature of language processing in the brain? Are the mechanisms that underlie perseverations necessarily tied to language in some way or are they common to other cognitive domains? This article examines what insights connectionist modeling can provide into these deeper questions.

Connectionist models are composed of relatively simple, neuron-like processing units that engage in parallel interactions through weighted connections. Units can be organized into groups that represent different types of information to be associated, such as acoustic, phonological, or semantic, within the domain of language. Connectionist models of cognitive processes have effectively addressed empirical results from a wide variety of different cognitive domains, including visual perception and attention,[13,14] reading and language,[15–17] semantic processing,[18–21] learning and memory,[22–24] working memory and cognitive control,[25,26] and routine sequential action.[27] One of the strengths of these models has been their ability to address not only behavioral results from neurologically intact adults but also basic behavioral impairments and patterns of errors

following neurological damage and behavioral changes during the course of normal development.[15,28–33]

The model most relevant for the current discussion is one proposed by Plaut and Shallice[34] to account for the large number of recurrent perseverations and semantic errors made by optic aphasic patients in visual object naming. The model was trained to identify visual objects by mapping information about an object's visual appearance to its corresponding semantic information. Learning in the model included short-term correlational weights that were strengthened each time an object was processed; these weights tended to bias activity in the model toward recently identified objects, producing perseverations under damage to the model's connections. The bulk of this article will be a review of the details of this model and a discussion of its implications for our understanding of aphasic perseveration. Some limitations of the model in explaining patient variability and in addressing more recent experimental findings on aphasic perseveration[5,35] are then briefly discussed. We conclude by suggesting modifications to the model that might address these limitations, taken from our understanding of how neuromodulatory systems in the brain interact with learning processes.

MODEL OF NAMING ERRORS IN OPTIC APHASIA

Before discussing the details of the Plaut and Shallice model,[34] we must first consider briefly the neuropsychological pattern of the *optic aphasia* that motivated it. Optic aphasic patients characteristically have difficulty naming objects presented visually but are able to name from other sensory modalities, such as from verbal description or touch. Unlike patients with visual agnosia, they show relatively preserved comprehension from vision in that they are able to appropriately mime object use for items they are unable to name.[9,36] It is also difficult to attribute this spared comprehension entirely to object "affordances" (actions biased by the object shape) or preserved high-level visuostructural information[37] (although see Hillis and Caramazza[38] and Riddoch and

Humphreys[39]). Patients with optic aphasia produce predominantly semantic and perseverative errors in picture naming, along with a smaller number of pure visual and other errors. This is a very different pattern from that of patients with visual associative agnosia, who tend to produce visual errors in naming.[40]

In their study of optic aphasic patient JF, Lhermitte and Beauvois[9] conducted one of the most thorough characterizations of recurrent perseveration in naming. These authors drew a distinction between *horizontal* and *vertical* influences in naming errors, referring to an error's relationship to the current stimulus—be it semantic, visual, or unrelated—and its relationship to a previous stimulus or response, respectively. More than 50% of JF's errors in naming pictures were perseverations (i.e., showing a vertical influence), with most of them also showing semantic or combined visual and semantic horizontal influences. Those of JF's errors that showed only a horizontal influence also tended to be semantic or combined visual and semantic errors with fewer pure visual errors. It should be noted here that the terms horizontal and vertical should not be confused with notions such as "paradigmatic" or "syntagmatic," semantic relations that have been discussed by other researchers (and refer to similarity versus contiguity relations, respectively, among stimuli).

The Plaut and Shallice model was proposed to address this particular pattern of horizontal and vertical influences in impaired naming, a pattern that emerges in the model from its basic learning mechanisms and how these mechanisms interact with properties of visual and semantic representations.[34] Because patients with optic aphasia do not appear to have language impairments other than their impaired visual naming, the model focused on the visual recognition component of the naming task. The model touches on issues of language processing, mainly in its inclusion of semantic or comprehension processes.

Model Architecture

The architecture of the model that simulates the recognition of visual objects is shown in Figure 1. The overall organization of the model consists of several different groups of units: 44 visual, 40 intermediate, 86 semantic, and 40 cleanup units. These groups were sparsely connected to each other, with the visual units connecting forward to the intermediate units and the intermediate units connecting forward to the semantic units. The semantic units connect both forward to and backward from the cleanup units, allowing feedback or *recurrent* interactions.

Through training, the model learns to generate the appropriate pattern of semantic activity across the semantic units when input representing a visual object is presented to the visual units. Thus, prior to any damage, the
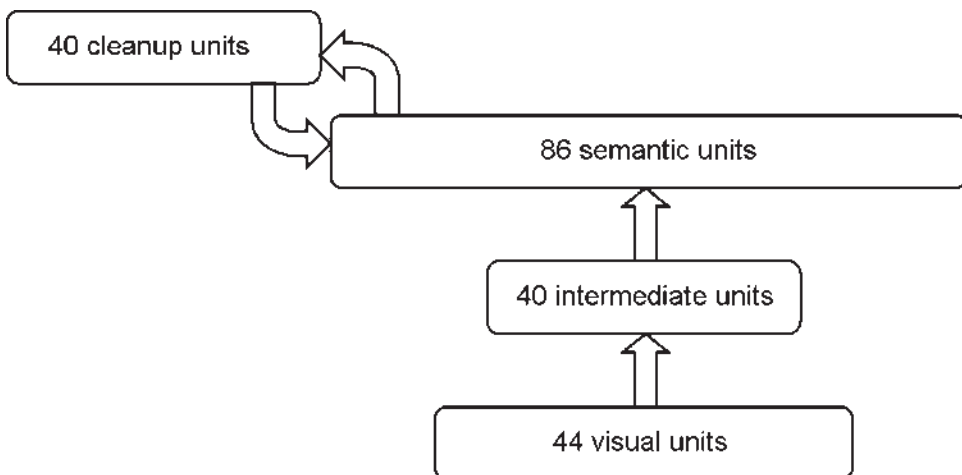

Figure 1 Architecture of the Plaut and Shallice[34] connectionist model.

model is set up to reflect the visual comprehension processes of non–brain-damaged, normal participants. The intermediate units serve to transform each visual input into an initial pattern of activity across the semantic units that then interact bidirectionally with the cleanup units to arrive at the final correct semantic pattern corresponding to the meaning of the visual object and the model's response. Artificial visual and semantic representations were generated for 40 common indoor objects from the categories of kitchen objects (e.g., *cup*), office objects (e.g., *pen*), furniture (e.g., *chair*), and tools (e.g., *saw*). Each visual pattern was distributed across 44 individual features that were intended to represent high-level visual information critical for object recognition. These patterns corresponded roughly to visual structural descriptions,[41,42] enhanced by information about color, texture, size, and additional general visual characteristics. Semantic patterns were distributed across 86 semantic features, 28 of which represented information about an object's visual semantics (e.g., abstract versions of the visual input features, including color, texture, size, shape, and other general visual characteristics), 2 representing the object's consistency (hard, soft), 8 representing the material from which it is made (metal, wood, cloth, and so on), 10 representing where it is found (home, office, kitchen, bedroom, and so on), 9 representing its general function (cooking, eating, leisure, aesthetic, and so on), 22 representing its specific function (chopping/cutting, measuring, container, and so on), and 7 representing its general action (use with one arm, use with two arms, and so on). To better appreciate the numerical calculations, understand that each visual and semantic feature took an "on" or "off" value of 1 or 0 for each object (see Appendix B of Plaut and Shallice[34] for a complete feature listing). Although these representations were clearly not exhaustive of all of the information people know about such objects, they were detailed enough to capture basic visual and semantic similarity relations among objects such that similar objects tended to share more of the same on and off values compared with unrelated objects and were thus related by similar numerical values.

Short-Term and Long-Term Learning

Learning is a critical feature of this and most other connectionist models. Rather than directly stipulating the values of weights on connections between groups of units, the model learns on its own the appropriate weights that ultimately allow it to map (to connect and relate) visual input to semantic output. Learning in the model between each pair of connected units $j \rightarrow i$ has two basic components: (1) standard long-term weights, w_{ij}, that are modified slowly over the course of training through *supervised* error-correcting learning and back propagation,[43,44] and (2) short-term weights, c_{ij}, that are modified through *unsupervised* correlational learning and that decay passively toward zero with the processing of each subsequent stimulus. The long-term and short-term weights w_{ij} and c_{ij} jointly influence the input to unit i at time t (denoted as $x_i^{(t)}$) from all units j that are connected to it through a simple weighted sum:

$$x_i^{(t)} = \sum_j s_j^{(t-1)}(w_{ij} + \gamma c_{ij}^{[n]}), \qquad (1)$$

where s_j is the activity state of sending unit j at time t–1 that ranges continuously from 0 up to 1, γ is a parameter that determines how strongly the short-term weight contributes to the total connection weight (set here to a small value of 0.1), and $c_{ij}^{[n]}$ refers to the current value of the short-term weight that is recalculated at the end of processing each stimulus n. Learning of the long-term weights w_{ij} proceeds in the following manner. Weights are initially set to small, random values at the beginning of training. A visual input pattern is presented to the visual units, and unit activities in subsequent groups of units are updated iteratively (changed progressively) as a function of their summed inputs $x_j^{(t)}$, allowing activity to spread along the weighted connections first to the intermediate units and then to semantic and cleanup units (see Appendix A of Plaut and Shallice[34] for more details). The semantic activity pattern actually produced by the input pattern at each time update is then compared with the desired or "teacher" semantic pattern (discussed in the previous section), and the resulting error signals are then used to make small adjustments to all of the long-term weights in the network to

reduce the error. In other words, the semantic patterns that were chosen by the researchers help to guide or constrain learning of the appropriate long-term weights in the model. Gradually, after many presentations of each training pattern, the model generates semantic unit activities to within 0.1 of the correct values at each unit for the all of the 40 objects.

In contrast to the learning of the long-term weights, the learning of the short-term weights c_{ij} depends on the recent correlations of unit activities: There is no supervision of what is actually produced compared with some target activity pattern. In this sense, the learning in the short-term weights is automatic and un-supervised. If s_i and s_j are the activity states of units i and j at the end of processing stimulus n, then the learning of short-term weight c_{ij} occurs in the following way:

$$c_{ij}^{[n+1]} = \lambda s_i' s_j' + (1 - \lambda) c_{ij}^{[n]}, \qquad (2)$$

where $s' = 2s - 1$, which realigns unit activities between -1 and $+1$ from 0 and 1 to allow agreeing unit activities of $1 = 1$ or $0 = 0$ to cause positive weight changes and disagree-ing unit activities of $0 \neq 1$ to cause negative weight changes (intermediate activities of 0.5 cause no change). λ is a parameter that deter-mines how much the unit states for the current stimulus n contribute to the new short-term weight relative to the weight's existing value $c_{ij}^{[n]}$. The value of λ used in the simulations was 0.5, implying that the weight changes due to a particular stimulus would decay rapidly toward 0 over two to three subsequent stimuli. This weight-change rule implements a simple form of correlation ($s_i' s_j'$) that tends to reinforce the current pattern of unit activity, "biasing" the network's current processing toward prior acti-vation states when the same units are activated again by the current stimulus. For example, if units i and j were both activated by the previous stimulus and the current stimulus reactivates one of the two units, the positive short-term weight c_{ij} will cause positive input to be sent to the other unit (see equation 1), increasing its likelihood of being active; similarly, if unit i was active during the previous stimulus but unit j was inactive, reactivation of unit i by the current stimulus will provide negative input to unit j

through weight c_{ij}, biasing it to be inactive again. We will see in the next sections that this bias toward prior activity states by the short-term weights is the critical factor that leads to recurrent perseverations under damage to the model's connections (damage analogous to the notion of "deafferentation" (Cohen and Dehaene[4]).

Simulating Brain Damage and Error Responses

As in other connectionist models of neuropsy-chological impairments,[29,31,32] brain damage in the model was simulated by removing a fraction of the connections between groups of units after the training phase (for example, removing 30% of the connections between intermediate and semantic units). The model's recognition per-formance under damage was tested in two-item sequences of *prime-target* pairs by presenting each of the objects as prime and fully crossing the primes with each object as target (for a total of $40 \times 40 = 1600$ prime-target pairs). Further-more, the presentation of each object as prime was redone for multiple samplings of damage to the model at each set of connections (visual $\rightarrow$ intermediate, intermediate $\rightarrow$ semantics, se-mantics $\rightarrow$ cleanup, cleanup $\rightarrow$ semantics) and over a range of damage severities (from 5 to 70% of connections at each location). The short-term weights were set to zero prior to the presentation of each prime in the prime-target pair, they were updated at the end of the prime presentation, and they were held fixed during the presentation of the target. The model was taken to have made a recognition response to the prime or target stimulus (be it correct or an error) if the resulting semantic unit states were sufficiently close to one of the trained semantic patterns, defined by a correla-tion or distance measurement across the seman-tic units. Otherwise, the model was taken to have produced an omission. If the model made an overt response, the response was considered correct if the generated semantic pattern was closest to the correct trained pattern, and it was considered an error if the generated pattern was closest to a different trained pattern than the correct one. Each error response to a target stimulus could then be classified with respect to

its *horizontal* relationship to the target (e.g., visual, semantic, combined visual and semantic, or unrelated), and its *vertical* relationship could be classified with respect to the prime response (e.g., identical = perseveration, semantically related to the prime but not identical = co-ordinate, or unrelated to the prime). It is important to reiterate that horizontal and vertical here are in the terminology of Lhermitte and Beauvois.[9] The terms essentially designate the temporal relationship between an error and a stimulus, with horizontal referring to an error's relationship to the correct response to the *current* stimulus and vertical referring to its relationship to the response to a *prior* stimulus (here, actually the immediately preceding prime stimulus). They are different from and should not be confused with notions of "paradigmatic" and "syntagmatic" semantic errors that have been used in some previous analyses of semantic errors.

Of the explicit error responses made by the model across all of the different damage locations and severities, more than 90% shared a semantic or combined visual and semantic horizontal relationship to the stimulus (e.g., responding "spoon" to the stimulus *fork*) whereas less than 8% were pure visual errors (e.g., responding "awl," a pointed tool for making holes in wood or leather, to the stimulus *fork*). Errors with a perseverative vertical relationship to the prime response accounted for approximately 29% of all errors, most of which also shared a semantic or visual and semantic relationship to the target (e.g., responding "spoon" to the target stimulus *fork* when the prime response was *spoon*). An additional 15% of the errors did not share an exact perseverative vertical relationship with the prime response but were instead semantically related to the prime (e.g., responding "fork" to the target stimulus *desk* when the prime response was *spoon*). This left slightly more than 50% of errors sharing no vertical relationship to the prime response at all, with most of these errors sharing a semantic or visual and semantic horizontal relationship to the target (e.g., responding "chair" to the target stimulus *desk* when the prime response was *spoon*) (see Bayles et al,[45] for a similar error typology in perseveration in Alzheimer's disease—EDS).

Why Does the Model Make Semantic and Perseverative Errors?

It appears then that, similarly to optic aphasic patients, many of the errors that the model makes in visual object identification following damage are semantic errors on the current stimulus or perseverations on the previous stimulus, or both. What are the mechanisms in the model that lead to this particular error pattern? A critical concept in understanding the functioning of this and other connectionist models with recurrent feedback connections is the notion of an *attractor*. When a visual pattern is presented to the visual units, activity in the semantic units changes over time. The initial pattern of semantic activity generated by the feed-forward pathway from the visual and intermediate units may be very different from the final pattern. The semantic units interact with the cleanup units to "clean up" the initially noisy or inaccurate semantic pattern. The final semantic states that result from the interactions with the cleanup units can be referred to as *attractors*, because the model will tend to be pulled into these states when the initial semantic states get close to them. The tendency to clean up noisy initial patterns into a known response is why the model tends to produce actual complete responses under damage rather than response blends or the semantic equivalent of neologisms. The range of initial semantic activities that will tend toward a final attractor semantic state are often referred to as the *basin of attraction* for that state. An idealized graphical depiction of this process is shown in Figure 2 for three stimuli: *chair*, *spoon*, and *fork*. This diagram depicts a geometric interpretation of the settling process, in which any given pattern of activity over a group of units corresponds to a particular point in a high-dimensional "state" space. Thus, visual and semantic patterns would correspond to points in spaces that have 44 and 86 dimensions, respectively (although Fig. 2 depicts only two dimensions for each). In each domain, the points for similar (overlapping) patterns share many coordinate values and hence are close to each other. For instance, stimuli such as *spoon* and *fork* are both visually and semantically similar to each other but dissimilar to *chair*. Notice that the points in vision and in
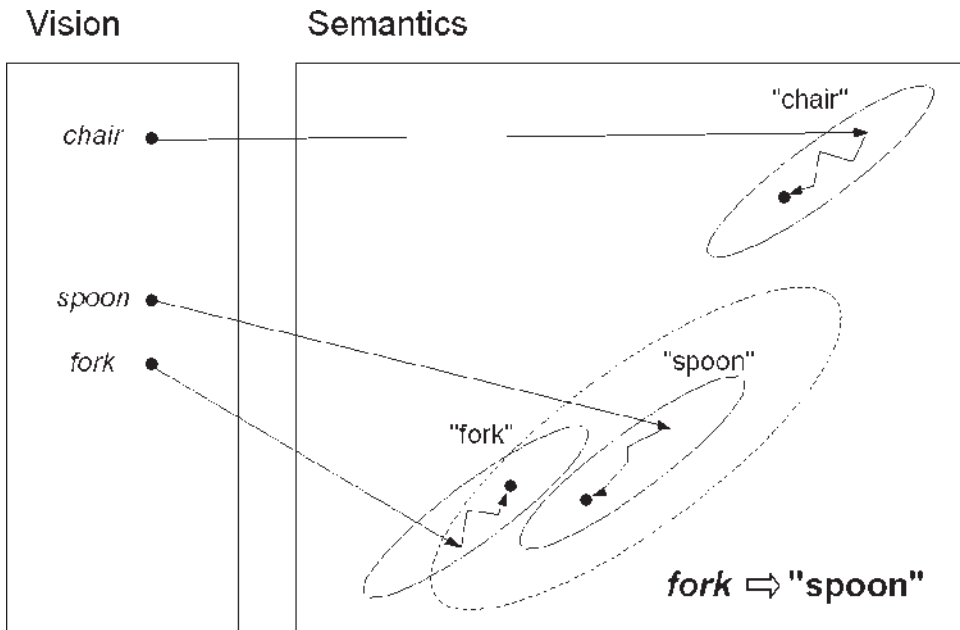
Figure 2 Geometric interpretation of the Plaut and Shallice[34] model settling into semantic attractors. Similar visual patterns (*spoon* and *fork*) tend to arrive at similar initial points in semantic space that are then progressively ''cleaned up'' through interactions between semantic and cleanup units to their final semantic states. Solid ovals in the semantic space define the basins of attraction for each object, and the dashed oval for *spoon* indicates the expansion of the normal basin due to the short-term correlational weights (see text for details).

semantics that correspond to *spoon* and *fork* are closer to each another than they are to *chair*. The arrows in the figure from vision to semantics show the initial activity points in semantic space generated by the feed-forward pathway from the visual and intermediate units. The semantic activity then moves along the jagged arrows because of the interactions with cleanup units to the final attractor state (shown by a dark, filled point) that correspond to the exact meaning of each visual object. The solid ovals represent the basins of attraction for each stimulus.

The long-term learning mechanisms of the model are responsible for the development of these arrows and attractor basins. Through learning, the model has to form the basins in such a way that it can correctly move from any point to any other with the initial push from the feed-forward pathway and the use of its semantic-cleanup interactions, despite a bias to remain in the previous activity state, due partially to the model's attractor dynamics and partially to the influence of the short-term correlational weights in reinforcing the last activity pattern. Correct performance requires the learning of

long-term weights that are strong enough to push the model out of its previous attractor state and into the attractor basin for the new stimulus, overcoming the influence of the short-term weights, which act like noise when stimuli randomly follow one another during training. Indeed, the influence of the short-term weights can be thought of as widening or deepening the basin of attraction temporarily for recent stimuli, shown in Figure 2 by the dashed oval for *spoon*. For these particular visual and semantic patterns, the learning pressures are different than they would be if this model were trained to recognize visual words, as was the connectionist attractor network studied by Hinton and Shallice.[29] The reason is that visually similar objects tend to be semantically similar too, such as *spoon* and *fork*, whereas the relationship between visual and semantic similarity is relatively arbitrary and unsystematic for visual words and their meanings.[34] This means that visual patterns representing similar visual objects do not need to be separated into very different initial semantic patterns by the feed-forward pathway; they can be relayed with less transformation. Similar visual objects will tend

to project to similar initial points in semantic space, as shown in Figure 2 for *spoon* and *fork*, and their respective attractor basins will tend to be close to one another compared with unrelated objects. In the event that two particular objects are visually similar but semantically very different, learning in the feed-forward pathway will tend to separate the visual activity patterns relatively early on in the pathway by developing strong weights from the visual units that are distinctive for the two objects (i.e., units that have different activity states), because this small number of units will have to override all of the shared visual feature information that is normally useful at determining the semantics of the objects.

When connections in the feed-forward pathway are removed to simulate brain damage, errors predominantly share a semantic or visual and semantic horizontal relationship to the target stimulus because the attractor basins for objects that are both visually and semantically similar are close together. The effect of damage is to distort by some amount the initial pattern of semantic activity for an object, potentially allowing it to fall in a nearby attractor basin that will be cleaned up to the exact meaning of a semantically similar or a visually and semantically similar object. Thus, the model will tend to produce semantic or combined visual and semantic errors. It will be much less likely to produce semantically unrelated or pure visual errors because these attractor basins are much further away from the correct basin than are ones corresponding to semantic associates. When connections between semantics and cleanup units are damaged, the model produces fewer explicit errors overall and more omissions, because these are the connections that implement the attractor dynamics and allow the model to arrive at exact object meanings. The explicit errors that the model does produce under these circumstances similarly tend to be semantic or combined visual and semantic errors.

Damage leads the model to produce perseverations on the response to the prime stimulus for a couple of reasons. The first and main one is that the short-term correlational weights effectively lead to a wider and deeper basin of attraction for the previous attractor state. This makes it difficult to leave the previous attractor state, particularly when the current stimulus is semantically or visually and semantically related to the prime. If the prime stimulus was *spoon* and the current target is *fork*, the basin of attraction for *fork* may now overlap partially with the enlarged basin for *spoon* because of the short-term weights (shown in Fig. 2 by the dashed oval), leading the model to return to the attractor for *spoon* again. As in the case of purely horizontal errors, perseverations will also tend to share a horizontal semantic relationship with the current target because of the proximity of attractor basins for semantic associates. Phrased more directly in terms of unit activity, when the current stimulus shares many of the same active semantic units with the previous stimulus (as in the case of semantic associates), the short-term correlational weights from these shared features start to reactivate units from the previous stimulus that should be off for the current stimulus and start to turn units off that should be on (see equation 2 earlier), leading the additional interactions between semantic and cleanup units to return the model to the previous semantic state. A second reason that the model may produce perseverations on the immediately preceding response is simply that it is less able to push out of its previous attractor state with weakened input resulting from damage to the feed-forward pathway. However, if this were the only reason that the model perseverated, then it would be unable to produce truly *recurrent* perseverations, those occurring after intervening trials and responses. Although the model was only assessed in two trial sequences of prime and target stimuli, the slowly decaying property of the short-term weight values (see equation 2) across subsequent stimuli would permit it to show perseverations after a small number of intervening stimuli, with perseverations becoming less likely with each intervening stimulus (matching empirical characteristics of recurrent perseveration[4,5]).

IMPLICATIONS AND LIMITATIONS OF THE CURRENT MODEL

Although the Plaut and Shallice model[34] is a model of visual recognition and only touches

directly on issues of language processing through its inclusion of semantic processing, it can account for many of the documented characteristics of aphasic perseveration. Its decaying short-term correlational weights will allow it to produce recurrent perseverations following a small number of intervening stimuli with fewer perseverations across longer delays.[4,5] It can also produce perseverations that share a horizontal relationship with the current target stimulus such as semantic ones,[3,10] and its tendency to produce perseverations will be influenced by factors in training or testing, such as stimulus repetition and lexical frequency.[5,15,35] These abilities to address characteristics of aphasic perseveration imply that similar mechanisms of learning, distributed representations, and attractor dynamics may underlie normal language processing. In other words, recurrent perseverations in aphasia may not reflect domain-specific language processes but instead reflect domain-general learning mechanisms that apply in both vision and language alike. This is consistent with the general connectionist approach to understanding cognition[46] that has attempted to show how a small set of domain-general computational principles can account for the richness of empirical data from a variety of different cognitive domains, including visual perception, attention, reading, language, memory, semantic memory, and working memory.

However, the Plaut and Shallice model[34] in its current form has a couple of major limitations that might undermine these conclusions. The first is that it fails to explain why some patients perseverate more than others. All locations of damage in the model produce a similarly high proportion of perseverations, roughly 30 to 40% of all explicit error responses. Although some patients exhibit rates of perseveration this high, such as the optic aphasic patient JF[9] or the aphasic patients EB[5] and CJ,[35] most patients with language impairments perseverate much less markedly. For example, tasks like picture naming elicit average perseveration rates well under 5% of total errors across patients from different aphasic categories.[47] It is unlikely that this patient variability is explained solely by severity of impairment. The second is that it is unable to

explain more recent empirical findings on aphasic perseveration,[5,35] such as the demonstration that perseverative responses following intervening stimuli can be unrelated to their target stimuli, instead reflecting the earlier sequential and temporal proximity of the same stimulus and response (i.e., if the response "fork" was given the trial before or after the stimulus *chair* on a previous occasion, the stimulus *chair* might later elicit the response "fork" again). The model is not able to form associations between sequentially presented stimuli or responses because the short-term correlational weights are updated only at the end of stimulus processing, after any hint of the prior semantic state has been pushed out by the processing of the current stimulus. It should be noted that these same limitations also apply to existing *priming* theories of perseveration, which explain perseverations as a failure of the current stimulus to override intact facilitatory mechanisms that lead to behavioral *priming* effects in normal subjects.[4,10] Indeed, the Plaut and Shallice model[34] is really a particular form of priming theory for which learning by the short-term and long-term weights will lead to the same stimulus being identified more rapidly and accurately after stimulus repetition. So it appears that, although connectionist models have the potential to provide deep insight into the mechanisms of aphasic perseveration, they may also have something to learn from their shortcomings in accounting for the entire range of characteristics.

Gotts and colleagues[5] outlined a remedy to both of these limitations in appealing to the possible neurophysiological and neurochemical bases of recurrent perseveration (see also McNamara and Albert[48]—EDS.). Several researchers have suggested previously that recurrent perseverations result from neuromodulatory deficits and low levels of acetylcholine.[49,50] Studies of the functional role of acetylcholine in the brain suggest that it serves to modulate the dynamics of cortical processing and learning, making cells more sensitive to bottom-up sensory signals by suppressing feedback or "recurrent" signals (see Haselmo[51] for a review). Under a cholinergic deficit, perseverations will be produced because cells are less sensitive to bottom-up sensory signals, making it harder for processing of the current

stimulus to override persistent neural activity that is enhanced by stronger recurrent feedback. Based on this view, the reason that some patients might perseverate more than others is that their brain damage may have affected subcortical cholinergic fibers that provide the brain with acetylcholine. Other patients might produce perseverative errors at a much lower rate because of a relative sparing of their neuromodulatory afferents. It is interesting to note on this point that the patients mentioned earlier who perseverated at high rates (such as JF and EB) had white matter damage that could have affected their cholinergic pathways (see Selden et al[52] for a review of the anatomy of cholinergic pathways). It is also possible to explain temporal or sequential effects of stimulus presentation on perseverations through abnormal learning that might occur under a neuromodulatory deficit. When feedback signals are strong at lower levels of acetylcholine, neural activity will effectively behave like the attractor dynamics exhibited by the Plaut and Shallice model.[34] As each new stimulus is presented, it will have to drive neural activity out of the previous state and into the correct new one. If the neural representations of two stimuli are coactive simultaneously as the new stimulus drives the old one out, rapid correlational learning among the active cells throughout this transition might allow the formation of inappropriate associations between sequentially presented stimuli (as in the *fork* and *chair* example in the previous paragraph). When one of the stimuli is presented again later, it might reactivate the representation of the other stimulus, producing a perseveration. This is not to say that sequential or temporal contiguity effects in learning are entirely abnormal. Indeed, the automatic learning of temporal contiguity is reflected in normal associative priming effects[53,54] (e.g., identifying *butter* can prime *knife*) and is undoubtedly critical for normal language and sequence learning. Nevertheless, neuromodulatory deficits might explain the marked and intrusive presence of such effects in some patients.

What modifications would be needed to the Plaut and Shallice model[34] to implement these properties of neuromodulation? First, it would be necessary to specify more about the relationship between connectionist models and real neural processes. A recent model by Gotts and Plaut[55] serves as a reasonable starting point. This model used a basic relationship between connectionist models and biophysical models of neural firing rate activity to suggest ways in which connectionist models can be made to incorporate neurophysiological and neuromodulatory mechanisms. Each group of units in the Plaut and Shallice model[34] would represent neural activity in anatomically distinct cortical regions that are functionally specialized for processing different types of information (e.g., the semantic units might represent neural activity in anterior, inferior temporal lobes that encodes semantic knowledge). The suppressive effect of acetylcholine then might correspond to a process that shuts down or suppresses interactions between the semantic and cleanup units that implement the model's attractor dynamics. Under a deficit of acetylcholine, attractor dynamics between the semantic and cleanup units would be much stronger, occasionally dominating the visual input from the feed-forward pathway. To account for the sequential effects of stimulus presentation on perseveration, short-term weights would have to be modified, not just at the end of stimulus processing but throughout processing. This would allow the short-term weights to behave more like real activity-dependent neural plasticity mechanisms (see Nelson et al[56] for a recent review) and would permit units activated by the current stimulus to form associations with units that were activated by the previous one, allowing perseverations to show sequential or temporal contingencies.

Importantly, the incorporation of neural principles such as neuromodulation and how it interacts with learning would not undermine the model's basic explanation of recurrent perseveration. These errors would still result from mechanisms of learning, distributed representations, and attractor dynamics. Instead, it would raise new questions about the impact of neuromodulatory mechanisms in language processing. How do these mechanisms shape the learning of representations in language and other domains? As we explore further the workings of connectionist models and bring them more into alignment with our understanding of

neural processes, they may provide useful revelations into these questions, too.

REFERENCES

1. Hudson AJ. Perseveration. Brain 1968;91:571–582
2. Sandson J, Albert ML. Varieties of perseveration. Neuropsychologia 1984;22:715–732
3. Albert ML, Sandson J. Perseveration in aphasia. Cortex 1986;22:103–115
4. Cohen L, Dehaene S. Competition between past and present: assessment and interpretation of verbal perseverations. Brain 1998;121:1641–1659
5. Gotts SJ, Incisa della Rocchetta A, Cipolotti L. Mechanisms underlying perseveration in aphasia: evidence from a single case study. Neuropsychologia 2002;40:1930–1947
6. Buckingham HW, Whitaker H, Whitaker HA. On linguistic perseveration. In: Whitaker HA, Whitaker H, eds. Studies of Neurolinguistics, Vol. 4. New York: Academic Press; 1979:329–352
7. Halpern H. Effect of stimulus variables on verbal perseveration of aphasic subjects. Percept Mot Skills 1965;20:421–429
8. Santo Pietro MJ, Rigrodsky S. The effects of temporal and semantic conditions on the occurrence of the error response of perseveration in adult aphasics. J Speech Hear Res 1982;25:184–192
9. Lhermitte F, Beauvois MF. A visual-speech disconnexion syndrome: report of a case with optic aphasia, agnosic alexia and colour agnosia. Brain 1973;96:695–714
10. Martin N, Roach A, Brecher A, Lowery J. Lexical retrieval mechanisms underlying whole-word perseveration errors in anomic aphasia. Aphasiology 1998;12:319–333
11. Papagno C, Basso A. Perseveration in two aphasic patients. Cortex 1996;32:67–82
12. Basso A. Perseveration or the Tower of Babel. Semin Speech Lang 2004;25:375–390
13. Behrmann M, Zemel RS, Mozer MC. Object-based attention and occlusion: Evidence from normal participants and a computational model. J Exp Psychol Hum Percept Perform 1998;24:1011–1036
14. Vecera SP, O'Reilly RC. Figure-ground organization and object recognition processes: an interactive account. J Exp Psychol Hum Percept Perform 1998;24:441–462
15. Plaut DC, McClelland JL, Seidenberg MS, Patterson K. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol Rev 1996;103:56–115
16. Rohde DLT, Plaut DC. Language acquisition in the absence of explicit negative evidence: how important is starting small? Cognition 1999;72:67–109
17. Dell GS, Schwartz MF, Martin N, Saffran EM, Gagnon DA. Lexical access in aphasic and nonaphasic speakers. Psychol Rev 1997;104:801–838
18. Farah MJ, McClelland JL. A computational model of semantic memory impairment: Modality-specificity and emergent category-specificity. J Exp Psychol Gen 1991;120:339–357
19. McClelland JL, Rogers TT. The parallel distributed processing approach to semantic cognition. Nat Rev Neurosci 2003;4:310–322
20. McRae K, de Sa VR, Seidenberg MS. On the nature and scope of featural representations of word meaning. J Exp Psychol Gen 1997;126:99–130
21. Plaut DC. Graded modality-specific specialization in semantics: a computational account of optic aphasia. Cogn Neuropsychol 2002;19:603–639
22. McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 1995;102:419–457
23. Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. Psychol Rev 2003;110:611–646
24. Stark CEL, McClelland JL. Repetition priming of words, pseudowords, and nonwords. J Exp Psychol Learn Mem Cogn 2000;26:945–972
25. Cohen JD, Dunbar K, McClelland JL. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychol Rev 1990;97:332–361
26. O'Reilly RC, Noelle DC, Braver TS, Cohen JD. Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. Cereb Cortex 2002;12:246–257
27. Botvinik M, Plaut DC. Doing without schema hierarchies: a recurrent connectionist approach to normal and impaired routine sequential action. Psychol Rev 2004;111:395–429
28. Cohen JD, Servan-Schreiber D. Context, cortex, and dopamine: a connectionist approach to

behavior and biology in schizophrenia. Psychol Rev 1992;99:45–77

29. Hinton GE, Shallice T. Lesioning an attractor network: investigations of acquired dyslexia. Psychol Rev 1991;98:74–95

30. Munakata Y, McClelland JL, Johnson MH, Siegler RS. Rethinking infant knowledge: toward an adaptive process account of successes and failures in object permanence tasks. Psychol Rev 1997;104: 686–713

31. Plaut DC, Shallice T. Deep dyslexia: a case study of connectionist neuropsychology. Cogn Neuropsychol 1993;10:377–500

32. Rogers TT, Lambon Ralph MA, Garrard P, et al. Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychol Rev 2004;111:205–235

33. Rogers TT, McClelland JL. A parallel distributed processing approach to semantic cognition: applications to conceptual development. In: Gershkoff-Stowe L, Rakison D, eds. Building Object Categories in Developmental Time. Hillsdale, NJ: Erlbaum; 2005 (in press)

34. Plaut DC, Shallice T. Perseverative and semantic influences on visual object naming errors in optic aphasia: a connectionist account. J Cogn Neurosci 1993;5:89–117

35. Hirsh KW. Perseveration and activation in aphasic speech production. Cogn Neuropsychol 1998;15: 377–388

36. Gil R, Pluchon C, Toullat G, et al. Visuoverbal disconnection (optical aphasia) for objects, pictures, colors and faces with abstractive alexia [in French]. Neuropsychologia 1985;23:333–349

37. Shallice T. The language-to-object perception interface: evidence from Neuropsychology. In: Bloom P, Peterson MA, Nadel L, Garrett MF, eds. Language and Space. Cambridge, MA: The MIT Press; 1995:531–552

38. Hillis AE, Caramazza A. Cognitive and neural mechanisms underlying visual and semantic processing: Implications from "optic aphasia." J Cogn Neurosci 1995;7:457–478

39. Riddoch MJ, Humphreys GW. Visual object processing in optic aphasia: a case of semantic access agnosia. Cogn Neuropsychol 1987;4:131–185

40. Iorio L, Falanga A, Fragassi NA, Grossi D. Visual associative agnosia and optic aphasia: a single case study and review of the syndromes. Cortex 1992; 28:23–37

41. Marr D, Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B Biol Sci 1978;200:269–294

42. Palmer SE. Hierarchical structure in perceptual representation. Cognit Psychol 1977;9:441–474

43. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–536

44. Williams RJ, Peng J. An efficient gradient-based algorithm for on-line training of recurrent network trajectories. Neural Comput 1990;2:490–501

45. Bayles KA, Tomoeda K, McKnight P, Helm-Estabrooks N, Hawley J. Verbal perseveration in individuals with Alzheimer's disease. Semin Speech Lang 2004;25:335–348

46. Rumelhart DE, McClelland JL, PDP Research Group. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. I. Cambridge, MA: MIT Press; 1986

47. Kohn SE, Goodglass H. Picture naming in aphasia. Brain Lang 1985;24:266–283

48. McNamara P, Albert ML. Neuropharmacology of verbal perseveration. Semin Speech Lang 2004;25: 309–322

49. Fuld PA, Katzman R, Davies P, Terry RD. Intrusions as a sign of Alzheimer's dementia: chemical and pathological verification. Ann Neurol 1982;11:155–159

50. Sandson J, Albert ML. Perseveration in behavioral neurology. Neurology 1987;37:1736–1741

51. Hasselmo ME. Neuromodulation and cortical function: modeling the physiological basis of behavior. Behav Brain Res 1995;67:1–27

52. Selden NR, Gitelman DR, Salamon-Murayama N, Parrish TB, Mesulam MM. Trajectories of cholinergic pathways within the cerebral hemispheres of the human brain. Brain 1998;121:2249–2257

53. Shelton JR, Martin RC. How semantic is automatic semantic priming? J Exp Psychol Learn Mem Cogn 1992;18:1191–1210

54. Neely JH. Semantic priming effects in visual word recognition: a selective review of current findings and theories. In: Besner D, Humphreys GW, eds. Basic Processes in Reading. Hillsdale, NJ: Lawrence Erlbaum Associates; 1991:264–336

55. Gotts SJ, Plaut DC. The impact of synaptic depression following brain damage: a connectionist account of "access" and "degraded-store" semantic impairments. Cogn Affect Behav Neurosci 2002; 2:187–213

56. Nelson SB, Sjostrom PJ, Turrigiano GG. Rate and timing in cortical synaptic plasticity. Philos Trans R Soc Lond B Biol Sci 2002;357:1851–1857